文章编号:1671-4229(2022)02-0042-18

视觉域泛化技术及研究进展

徐 海,谢洪涛*,张勇东
(中国科学技术大学 信息科学技术学院,安徽 合肥 230027)

摘 要:近年来,机器学习理论和深度学习算法在计算机视觉领域发展迅速,并且在目标检测、语义分割、动作识别等任务场景中得到广泛应用。然而,实际部署中模型效果往往依赖于训练域和测试域服从独立同分布这一假设,受域偏移(Domain shift)现象影响严重。域偏移(即目标域数据分布与训练域不一致)对模型的泛化性提出了巨大挑战,使得域泛化(Domain generalization)技术成为计算机视觉领域一个重要的研究方向。域泛化研究如何在单一或者多个源域上进行模型训练,使其能够在具有不同数据分布的未知目标域上保持良好的泛化性,为模型应用提供了重要的保障。文章对近年来计算机视觉领域中域泛化研究具有代表性的论文进行梳理和总结,概述视觉域泛化技术及其研究进展。首先对域泛化的任务定义、任务特点和研究思想进行详细阐述;其次,遵循域泛化研究思路,从增广数据空间、优化模型求解和减小域间差异3个大方向分类总结域泛化领域的最新研究成果;随后介绍了域泛化技术在计算机视觉任务中的应用以及已公开的大规模数据集;最后讨论了域泛化研究领域未来可能的研究方向。

关键词:域偏移;域泛化;模型鲁棒性;深度学习;人工智能
中图分类号:TP 37 **文献标志码**:A

Review of domain generalization in vision

XU Hai, *XIE Hong-tao**, *ZHANG Yong-dong*
(School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China)

Abstract: Recent years have witnessed a rapid development of machine learning theories and deep learning algorithms in the field of computer vision, which have been widely applied in task scenarios such as object detection, semantic segmentation and action recognition. However, model performances in practical deployment always rely on the assumption that the training and testing data are identically and independently distributed, and prone to be affected by domain shift. Domain shift, namely distribution shift between the training domain and the test domain, poses a great challenge to the model generalization, thereby making domain generalization an important research topic in computer vision. Domain generalization aims to train a model on a single or multiple domains and maintain good generalization on unknown target domains with different distributions, which provides important guarantees for model deployments. This paper clearly elaborates the domain generalization technique in the computer vision field and its research progress through combing and summarizing representative researches over the past decade. Firstly, the paper makes a comprehensive introduction of domain generalization from three aspects, i. e., task formulation, task characteristics and research ideas. Then, the paper catergorizes existing approaches into three groups according to the main research ideas to deal with domain shift, namely, data augmentation, model optimization, and domain gap mitigation. Sub-

基金项目:国家自然科学基金创新研究群体资助项目(62121002);国家自然科学基金优秀青年科学基金资助项目(62022076);国家自然科学基金联合基金重点资助项目(U1936210)

作者简介:徐海(1993—),男,博士研究生. E-mail:keda2010@mail.ustc.edu.cn
* 通信作者. E-mail:htxie@ustc.edu.cn

引文格式:徐海,谢洪涛,张勇东. 视觉域泛化技术及研究进展[J]. 广州大学学报(自然科学版),2022,21(2):42-59.

sequently, the applications of domain generalization techniques in computer vision and related large-scale public datasets are introduced. Finally, the paper discusses problems needed to be solved in the current domain generalization field and possible future research directions.

Key words: domain shift; domain generalization; model robustness; deep learning; artificial intelligence

当前,机器学习在自然语言处理、计算机视觉和医疗健康等领域的成功推动着人工智能从纯粹的学术研究向产业落地转变[1-2]。传统机器学习算法遵循训练域和测试域服从独立同分布(independent and identical distribution, i.i.d)这一基本假设,然而由于数据获取设备和条件的差异,现实场景中数据有可能呈现出和训练集不同的分布。模型训练域和测试域分布不一致的现象在机器学习中被称为域偏移。域偏移现象在计算机视觉领域中广泛存在,物体分类任务中图像风格的差异、行人重识别任务中拍摄角度和相机的变化、医学影像中成像设备或参数的不同、自动驾驶任务中数据采集时天气、光照条件的区别,都可能会引起数据域的分布发生偏移。

与人脑认知方式相比,机器学习对数据域的分布变化更为敏感。例如小孩能够很容易地在动物园中准确认出只在动画片中见过的动物,但这对于仅使用卡通数据进行训练的分类模型来说却很困难。研究表明[3-4],当训练集和测试集分布存在明显差异时,机器学习模型性能容易出现大幅度的下降。因此,域偏移问题严重阻碍了模型进行大规模的部署应用,现实场景中数据的多变性和不可预知性对模型的域泛化能力提出了巨大的挑战[5]。

一般来说,域泛化旨在研究如何在单一或者多个相似但分布不同的源域(Source domain)上进行模型训练,使其能在具有不同分布的未知目标域(Target domain)上也能保持良好性能。

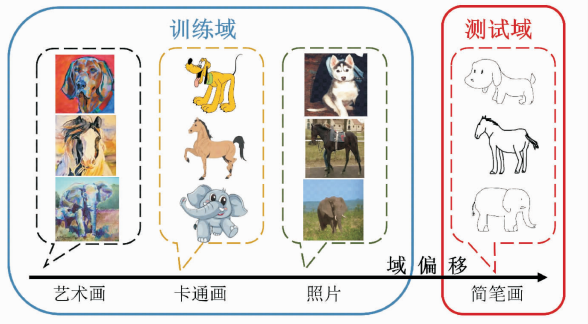

图 1　PACS[6]数据集上域泛化任务示例

Fig. 1　Example of domain generalization on PACS

图 1 示例中,PACS[6]数据集上域泛化任务要求在包含艺术画、卡通画以及照片 3 个源域的数据集上训练的模型,在目标域简笔画数据上也能取得良好的泛化性能。

域泛化技术对模型泛化性研究具有重大意义,近年来引起了科研人员广泛的研究兴趣[7-8]。域泛化任务的挑战在于如何克服域偏移问题,从源域数据中学习到更一般化的、迁移性更好的特征表达。随着深度学习技术的发展进步及其在域泛化相关研究领域(如域适应、迁移学习和元学习)内取得的显著成果,以深度学习技术为基础涌现了一系列的域泛化研究工作,并取得了优异的模型泛化性能,对学术界和工业界产生了积极影响。

尽管域泛化研究在计算机视觉领域取得了一定的成果,但作为相对新兴且小众的研究方向,目前关于此方向的综述还不全面[9-10]。本文旨在通过对近年来计算机视觉领域内具有代表性的域泛化研究工作进行梳理和总结,对域泛化的任务定义、任务特点和研究思路进行详细概述,从而有助于研究人员快速了解该领域。区别于文献[9-10]综述,本文结合了传统模型泛化理论和基于域适应的域泛化理论,从影响模型泛化性能的因素中提炼出域泛化研究思路,并以其为线索,对应地从增广数据空间、优化模型求解、减小域间差异 3 个大方向分类总结域泛化领域的最新研究成果,并重点阐述了最新的研究进展,如基于频域空间的增广方法。此外,还介绍了域泛化技术在计算机视觉领域中的最新应用情况所对应的公开数据集,如深度伪造检测任务。最后,讨论了域泛化领域未来可能的研究方向。

1　域泛化概述

1.1　模型泛化性

模型泛化性指的是在给定数据集上训练的模型对未观测样本的适应能力,为模型的部署应用提供重要的支撑作用。随着计算资源和深度学习技术的发展迭代,深度神经网络(Deep Neural Networks, DNNs)在图像分类、医疗健康、自然语言处理和计算机视觉等领域中取

得了媲美乃至超越人类的表现[11-12],但其泛化能力与人类仍存在一定的距离。

经典的泛化理论研究通常假设训练样本和测试样本来自同一概率分布,通过测试误差来评价模型的泛化能力,然而其无法应对目标域分布和训练域存在域偏移的情况。为了揭示DNNs和人脑认知在目标分类任务上的泛化性差异,Geirhos等[13]通过12种不同的图像畸变方法模拟不同的测试域,并统计人类和机器的分类准确率。结果表明,DNNs和人类在认知方式上仍存在显著差别,当潜在的目标域数据不再和源域数据服从独立同分布这一假设时,DNNs泛化性能会明显下降。为了提升模型在多变的现实场景下的适用性,比如识别模型能够不受图像风格的影响,自动驾驶系统能够对不同天气条件下的数据保持鲁棒性[14],同时增强对深度学习的理解,研究模型域泛化具有重要意义。

1.2 域泛化定义

域泛化研究旨在单一或者多个相似,但分布不同的源域上进行模型训练,使其能够在未知的不同分布下的目标域上也能保持良好的泛化性能。

1.2.1 符号表示

给定 d 维样本空间 $X\subset \mathrm{R}^d$,标签空间 $Y\subset \mathrm{R}$,数据域 $D:=\{D_i\}_{i=1}^{I}$ 用来表示 I 个训练(源)域的集合,其中,第 i 个域 $D_i=\{(x_j^i,y_j^i)\}_{j=1}^{N_i}\sim P_{XY}^i$ 表示服从联合分布 P_{XY}^i 的 N_i 个样本点 (x_j^i,y_j^i) 的集合,对于域泛化任务,源域之间的联合分布是不同的,即 $P_{XY}^i\neq P_{XY}^j$。同时,定义目标域 $T=\{x_i^T\}_{i=1}^{N_T}$,且目标域分布与源域分布也不同,$P_{XY}^T\neq P_{XY}^i$。

1.2.2 任务定义

域泛化,即给定 I 个具有不同数据分布的源域 $D_{train}:=\{D_i\}_{i=1}^{I}$,利用有标签的源域数据在假设函数空间上学习一个模型 $h(\cdot;\theta):X\mapsto Y(\theta\in\Theta$ 为模型参数),通过最小化源域上的损失 $\varepsilon_{D_{train}}(\theta)$ 使得模型 h 在未知的目标域 T 上预测误差 $\varepsilon_T(\theta)=\mathbb{E}_{(x,y)\in T}[\ell(h(x;\theta),y)]$ 最小。其中,$\ell:\mathrm{Y}\times\mathrm{Y}\to[0,\infty]$ 是损失函数,且有

$$\varepsilon_{D_{train}}(\theta)=\frac{1}{I}\sum_{i=1}^{I}\mathbb{E}_{x^i\sim D_i}[\ell(h(x^i;\theta),y^i)] \quad (1)$$

1.3 域泛化任务特点

域泛化任务定义示意图如图2所示,区别于传统泛化性研究,域泛化有如下特点:①训练集通常包含单个或者多个分布不同的源域;②目标域不可见且分布与源域不同。域泛化和相关研究领域具体对比如下:

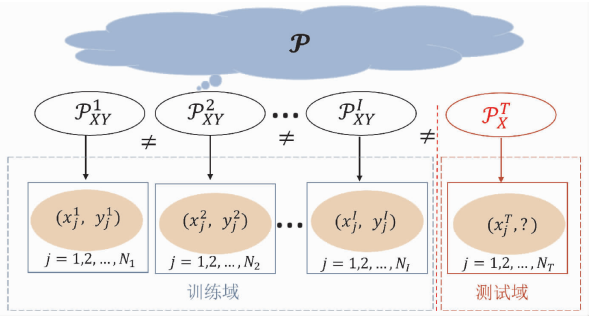

图2 域泛化任务定义示意图

Fig. 2 Illustration of domain generalization formulation

(1)监督学习(Supervised learning):与单源域的域泛化任务相似,监督学习仅在单一源域上进行模型训练,其模型泛化性严重依赖于源域和目标域独立同分布这一假设。经典泛化理论中关于模型复杂度和数据容量的论述对域泛化具有启发作用。

(2)多任务学习(Multi-task learning):模型并行地学习多个任务,以提升模型在其中某目标任务下的学习效率和预测准确性[15-16]。多任务学习之所以有效,是因为要求算法在相关任务上表现良好,并起到某种正则化效果,从而学到更一般化的特征。学习更一般化的特征对于域泛化具有借鉴作用。

(3)迁移学习(Transfer learning):旨在将源域上学习到的领域知识迁移到目标域中,以提升目标域任务的模型性能[17-19]。比较常见的迁移方式是用在源域上预训练好的模型对目标域上的模型进行初始化,然后在目标域上进行微调(Finetune),或者通过知识蒸馏的方式使模型能够学到源域的知识。与域泛化不同,迁移学习中目标域是可见的,且目标域任务可以与源域差异很大。

(4)元学习(Meta-learning):旨在让模型学会如何学习(Learning to learn),即利用先前任务中学习到的先验知识去指导新任务的学习,使其能够根据新任务进行快速调整[20-21]。域泛化任务中通过对源域进行划分构建元任务,可以隐式地在不同域之间添加更新方向一致性约束。与元学习不同的是,域泛化中目标域是不可见的。

(5)零样本学习(Zero-shot learning):基于可见源域上的数据和标签训练模型,并在未知的目标域上进行预测。零样本学习中未知目标域分布和源域不同之处主要体现在标签空间上,目标域数据都来自于未见过的标签类别。零样本学习旨在表征学习和度量学习基础上,借助属性或者文本信息实现知识的迁移[22]。

(6)域适应(Domain adaptation):解决域偏移问题中最直接的方法,即研究目标域可知的情况下,跨域泛化

性(Cross-domain generalization)问题。域适应可以看作是背景条件约束更严格的迁移学习问题,让在源域上训练的模型去适应目标域上的数据分布[23-24]。与域泛化相比,域适应下目标域数据是可获取的。作为与域泛化最相近的研究领域,域适应任务为跨域泛化性研究提供了理论基础。

域泛化和各研究领域的对比总结见表 1,尽管不同研究领域之间存在差异,其他领域(如域适应)的研究思想对域泛化存在广泛的借鉴意义。

表 1　域泛化与各研究领域对比

Table 1　Comparisons between domain generation and related tasks

项目	源域数目	域偏移	标签空间	目标域
监督学习	=1	×	一致	×
多任务学习	≥1	×	一致	×
迁移学习	≥1	√	不一致	√
元学习	>1	√	都可能	√
零样本学习	=1	√	不一致	×
域适应	≥1	√	都可能	√
域泛化	≥1	√	都可能	×

2　域泛化研究思想

2.1　传统泛化理论

传统泛化理论揭示了模型泛化误差与训练样本数量和模型复杂度之间的关系,对域泛化研究起到一定的指导作用。在经典的统计学习理论中,模型泛化性能可以用泛化误差进行度量。泛化误差表示为模型 h 在数据域 $(X,Y)\sim P_{XY}$ 上的期望风险,表达式如下:

$$R(h)=\mathbb{E}_{(X,Y)\sim P_{XY}}[\ell(h(X),Y)] \tag{2}$$

文献[25]证明了泛化误差 $R(h)$ 至少以 $1-\delta$ 的概率满足

$$R(h)\leqslant\hat{R}(h)+\xi(d,N,\delta) \tag{3}$$

式(3)中,$\hat{R}(h)$ 为模型 h 在训练集数据上的经验风险,$\xi(d,N,\delta)$ 与模型假设空间 $\mathcal{H}=\{h_1,h_2,\cdots,h_d\}$ 的容量 d 以及训练样本数目 N 有关,具体关系为

$$\xi(d,N,\delta)=\sqrt{\frac{1}{2N}\left(\log d+\log\frac{1}{\delta}\right)} \tag{4}$$

从式(4)可以得出,模型泛化误差上界由训练样本容量 N 和模型假设空间容量 d 共同约束,且具有如下性质:①当样本容量增加时,泛化误差上界趋于零,模型泛化性能越强;②模型假设空间容量越大,泛化误差上界越大,模型学习难度增加,泛化能力减弱。因此,传统基于模型容量的泛化理论认为对训练数据的过拟合是导致泛化性能差的重要原因,且符合奥卡姆剃刀原则的学习策略和正则化方式,有助于模型泛化能力的提升。

2.2　域泛化理论

和传统泛化理论中训练域和测试域独立同分布的假设不同,域泛化的目标是在具有不同分布的源域数据上学习通用的特征表达,并希望该特征表达也能应用于未见过的目标域数据。通常可以用目标域上的泛化误差去衡量该特征表达的泛化性能,然而域泛化任务中目标域是不可见的,目标域上的泛化误差无法直接度量。鉴于域泛化研究间接约束了不同源域之间的跨域泛化性,源域之间的域适应问题研究为域泛化提供了理论指导[26]。

考虑二分类任务上的域适应问题,源域和目标域分别用 $<D_S,f_s>$ 和 $<D_T,f_T>$ 表示,其中,D 表示域内数据分布,$f:X\mapsto[0,1]$ 为标签函数。在源域 D_S 上训练的模型用假设函数 $h:X\mapsto[0,1]$ 表示,则源域泛化误差 $\varepsilon_S(h,f_S;D_S)$ 用期望风险表示如下:

$$\varepsilon_S(h,f_S;D_S)=\mathbb{E}_{x\sim D_S}[|h(x)-f(x)|] \tag{5}$$

为简化表示,记 $\varepsilon_S(h)=\varepsilon_S(h,f_S;D_S)$,$\hat{\varepsilon}_S(h)$ 为源域上的经验风险估计。类似地,目标域上的误差用 $\varepsilon_T(h)$ 和 $\hat{\varepsilon}_T(h)$ 表示。域适应的目标是将源域 D_S 上训练的模型泛化到目标域 D_T 上,然而目标域标签函数 f_T 未知,无法直接度量泛化误差 $\varepsilon_T(h)$。文献[27]证明了可以通过 L^1 距离给出 $\varepsilon_T(h)$ 的上界,即

$$\varepsilon_T(h)\leqslant\varepsilon_S(h)+d_1(D_S,D_T)+\min\{\varepsilon_S(f_S,f_T),\varepsilon_T(f_S,f_T)\} \tag{6}$$

其中,$d_1(D_S,D_T)=2\sup_{B\subset X}|\mathrm{P}_{D_S}[B]-\mathrm{P}_{D_T}[B]|$ 是基于 L^1 范数的变分散度(Variation distance)。式(6)右边第三项表示源域和目标域标签函数的差异,在协变量偏移假设下可以忽略。然而,$d_1(D_S,D_T)$ 无法精确估计有限样本下的分布距离且求上确界的条件过于苛刻。实际上只需要关注对于假设函数 h 而言的域之间的距离,为了将域分布的距离与目标假设联系起来,文献[27]提出了 $\mathcal{H}$-距离($\mathcal{H}$-divergence):

$$d_{\mathcal{H}}(D_S,D_T)=2\sup_{h\in\mathcal{H}}|\mathrm{P}_{D_S}[I(h)]-\mathrm{P}_{D_T}[I(h)]| \tag{7}$$

式中,$I(h)=\{x|h(x)=1\}$ 表示与假设函数 h 相关的数据域的一个子集。要想让 $\mathcal{H}$-距离很小,需要在假设空间上训练的分类器 h 能够准确区分输入数据来自源域还是目标域,$d_{\mathcal{H}}$ 可以使用经验误差 $\hat{d}_{\mathcal{H}}$ 进行计算。

为了利用 $d_{\mathcal{H}}(D_S,D_T)$ 来建立 $\varepsilon_T(h)$ 和 $\varepsilon_S(h)$ 的关系,首先引入理想联合假设 h^*,满足

$$h^*=\mathrm{argmin}_{h\in\mathcal{H}}(\varepsilon_S(h)+\varepsilon_T(h)) \tag{8}$$

记 h^* 下理想联合预测误差 $\lambda=\varepsilon_S(h^*)+\varepsilon_T(h^*)$,其次

基于异或的思想定义对称假设空间$\mathcal{H}\Delta\mathcal{H}$:

$$g\in\mathcal{H}\Delta\mathcal{H}\Leftrightarrow g(x)=h(x)\oplus h'(x) \tag{9}$$

式中⊕表示异或操作,$\mathcal{H}\Delta\mathcal{H}$空间用来度量2个假设$h$和$h'$的不一致性,$D_S$和$D_T$在$\mathcal{H}\Delta\mathcal{H}$空间上的距离度量表示为

$$\begin{aligned} d_{\mathcal{H}\Delta\mathcal{H}}(D_S,D_T)&=2\sup_{h,h'\in\mathcal{H}}\left|\Pr_{D_S}[h(x)\neq h'(x)]-\right.\\ &\quad\left.\Pr_{D_T}[h(x)\neq h'(x)]\right|\\ &=2\sup_{h,h'\in\mathcal{H}}|\varepsilon_S(h,h')-\varepsilon_T(h,h')| \end{aligned} \tag{10}$$

由此可得

$$|\varepsilon_S(h,h')-\varepsilon_T(h,h')|\leqslant\frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}}(D_S,D_T) \tag{11}$$

且容易求得$d_{\mathcal{H}\Delta\mathcal{H}}(D_S,D_T)\leqslant 2d_{\mathcal{H}}(D_S,D_T)$,因此有

$$|\varepsilon_S(h,h')-\varepsilon_T(h,h')|\leqslant d_{\mathcal{H}}(D_S,D_T) \tag{12}$$

最终给出了目标域泛化误差的上确界:

$$\begin{aligned} \varepsilon_T(h)&\leqslant\varepsilon_T(h^*)+\varepsilon_T(h,h^*)\\ &\leqslant\varepsilon_T(h^*)+\varepsilon_S(h,h^*)+|\varepsilon_S(h,h^*)-\\ &\quad\varepsilon_T(h,h^*)|\\ &\leqslant\lambda+\varepsilon_S(h)+|\varepsilon_S(h,h^*)-\varepsilon_T(h,h^*)|\\ &\leqslant\varepsilon_S(h)+\lambda+d_{\mathcal{H}}(D_S,D_T) \end{aligned} \tag{13}$$

经过上述推导可知,目标域泛化误差的上界由3个因素界定,分别是源域误差、理想联合预测误差和域间距离。其中,源域误差和理想联合预测误差主要受模型假设空间影响,当模型结构固定时,源域和目标域在假设函数h下的分布距离是影响目标域泛化误差的关键因素。因此,通过特征空间过渡,使得在特征空间上目标域和源域无法区分,是保证跨域泛化性的重要手段。

2.3 域泛化研究思路

2.3.1 增广数据空间

基于传统泛化理论,域泛化研究的第一种思路是增广数据样本空间(Data augmentation)。这一类方法的基本思想是在原有训练数据的基础上,通过数据增广技术产生更多的数据用于训练,降低模型过拟合的风险,从而提升模型的泛化性。目前,域泛化领域数据增广方法大致可以分为2大类:第一类是基于图像处理技术的数据增广,主要包括几何变换、颜色变换和图像融合等;第二类是基于深度学习方法的数据增广,主要涉及到图像风格转换、对抗样本、生成对抗网络及特征空间增广等多种技术。

2.3.2 优化模型求解

传统泛化理论认为,泛化误差与模型容量成正比,且模型容量可以简单地用模型参数多少来表示。然而在实践中,DNNs通常包含比训练样本更多的参数,对数据拟合能力更强,却表现出出色的泛化性能,继续增大模型的参数量,模型的泛化性能也不会变差。Zhang等[28]将DNNs的泛化能力一部分归因于模型的记忆能力,即模型容量足够大到可以记住所有训练数据。随后,Krueger等[29]实验发现,对于真实数据,DNNs用较少的参数获得较好的性能,而对于噪声则需要增加模型的容量。这表示网络不仅仅是简单的暴力记忆,而是从数据中学习某种模式。进一步,文献[30]总结了前面的工作,指出DNNs泛化性不仅与模型容量有关,优化策略以及数据本身都会对泛化性造成影响。

因此,域泛化研究的第二种思路是优化模型解空间,以降低模型复杂度。通常域泛化认为不同域之间联合分布不同是由协变量偏移(Covariate shift)引起的,即不同域的边缘分布$P_X^i\neq P_X^j$不同,而标签函数(后验分布)被认为是相同的$P_{Y|X}^i=P_{Y|X}^j$。协变量偏移会导致模型求解困难,增加模型复杂度,进而降低泛化性。因此,在模型容量大小固定的前提下,如何优化模型解空间,使得模型在泛化性更好的模型解子空间下求解,可以提升域泛化性能。最典型的降低模型复杂度的方式是符合奥卡姆剃刀原则的传统正则化方法,如Weight Decay。目前,域泛化领域主要围绕基于深度学习的标准化算法、集成学习和学习策略进行。比如基于元学习的思想优化求解过程,从而达到优化解空间的目的。

2.3.3 减小域间差异

域泛化研究中,由于目标域是不可见的,无法直接度量目标域的泛化误差,但是可以通过保证源域内的跨域泛化性隐式地提升域泛化性能。因此,在源域上学习一个通用的特征表达,使得不同域之间的差异变小,是域泛化研究的第三种思路。减小域间差异的基本思想是学习具有域不变性的特征表达(Domain-invariant representation),域不变性保证了特征对域偏移不敏感,因此能更好地泛化到不可见的目标域上。基于特征解耦(Feature disentanglement)的域泛化方法也被广泛研究,这类方法认为特征空间可以被解耦成域不变(Domain-invariant)特征和域特异(Domain-specific)特征2个部分。此外,学习数据更一般化的特征(Generic features),使模型不仅更关注数据语义信息而忽视域的特定偏差(Domain-specific bias),也能减小域间差异,使模型更容易泛化到目标域上。

3 域泛化方法

本节遵循域泛化研究思路,对近年来具有代表性的研究工作进行归类,如图3所示。

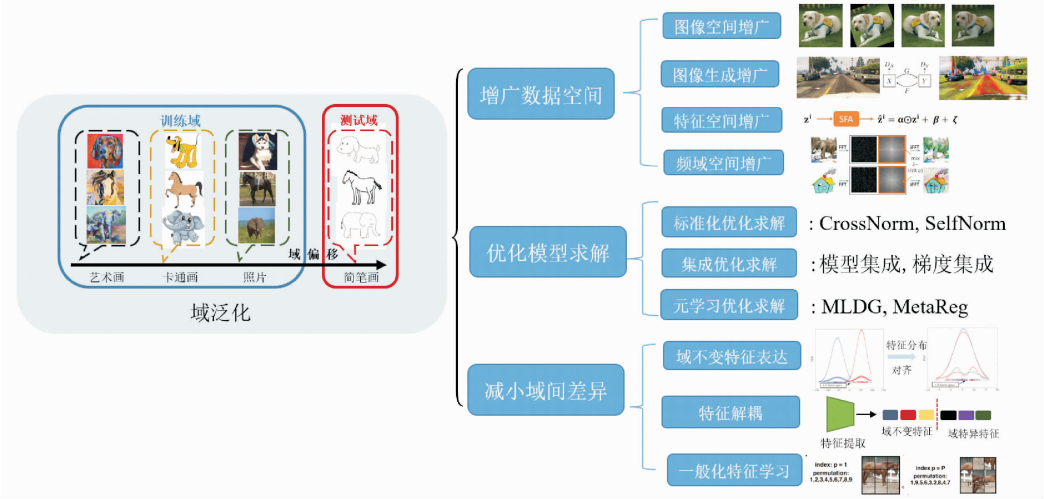

图 3　域泛化方法归类

Fig. 3　Overview of domain generalization methods

3.1　增广数据空间

随着深度学习技术和算力的发展,DNNs 模型参数呈指数增长,模型数据拟合能力得到前所未有的提升。为了降低深度学习模型过拟合的风险,能生成更多训练数据的数据增广技术是最直接有效的方式,从而被广泛研究用于提升模型的鲁棒性。同样地,在域泛化研究领域,数据增广技术可以用来模拟数据域的多样性,降低模型受特定域偏差的影响,以达到提升模型在目标域上的泛化效果。根据数据样本增广方式的不同,本文将基于数据增广的域泛化研究方法分成 4 个类别进行介绍。

3.1.1　基于图像空间的数据增广

此类数据增广方法基于传统的图像处理技术进行数据增强,主要包含几何变换、颜色变换、噪声注入和图像融合等方式。这些增广方式对于域偏移情形相对简单的域泛化任务非常有效,如源域图像差异主要体现在位置视角和亮度变化的手写数字识别任务中。根据域泛化任务中域偏移产生的原因不同,针对性地设计数据增强方法可以有效提升域泛化效果。在手写数字识别任务中,MNIST、MNIST-M、Rotated MNIST 等数据域的差异主要体现在字体、颜色、背景和旋转角度上,因此,可以采用几何变换和颜色变换相关的数据增广方式以模拟可能的目标域。而对于受天气变化影响最大的自动驾驶任务来说,选择图像对比度、亮度和曝光度等颜色变换进行数据增广更贴近潜在的域偏移情形。那么如何选择变换是最优的呢? 直观上认为与当前源域空间风格差异大的图像变换方法收益更大。Volpi 等[31] 设计了一套基于演进的搜索算法,迭代式搜索能使变换后的图像分布尽可能偏离原始图像分布的变换方法,并证明了该方法能有效提升模型泛化性。在人脸识别任务中,文献[32]选择了降低图像分辨率、添加遮挡和改变头部姿势这些增加识别难度的图像变换方法。

然而,传统的数据增强方法通常适用于单一数据域中的简单图像变化,无法处理源域间数据差异很大的情况,如包含明显图像风格变化的 PACS。

3.1.2　基于图像生成的数据增广

图像生成是域泛化领域进行数据增广的一种常见方式,通过生成大量的、多样化的数据来提高模型的泛化能力。最常见的图像生成方式是借助生成对抗网络(Generative Adversarial Network, GAN)、变分自编码(Variational Auto-Encoder,VAE)等生成模型产生不同风格的图像。Yue 等[33] 使用 cycleGAN[34] 将合成的数据映射到不同的风格空间以实现域随机化(Domain randomization),并对不同域下的图像添加一致性约束以学习域不变特征表达。Rahman 等[35] 使用 ComboGAN[36] 生成新的数据,然后应用最大均值差异[37] (Maximum Mean Discrepancy, MMD)度量最小化真实和生成图像之间的分布差异,以帮助学习到更通用的特征表示。在域泛化领域,梯度信息也被广泛研究用来图像生成。受对抗攻击启发,Volpi 等[38] 利用分类器梯度信息生成包含对抗噪声的对抗样本图像,并结合对抗训练提升模型泛化性;Qiao 等[39] 还使用了 WAE 以保留样本语义信息,并使其与源域具有最大差异。为了使模型学习到更多具有域不变性的特征,文献[40-41]在模型基础上设计了域分类器,利用域分类器的梯度信息设计扰动,使生成图像能够骗过域分类器。基于梯度信息生成的图像在视觉上与原图无差异,因此会被诟病无法模拟真实世界的域偏移情况。

3.1.3 基于特征空间的数据增广

考虑到依赖于图像空间的方法需要精心的增强设计,而且仅能够提供有限的增强数据多样性,同时基于图像生成的数据增广方式需要引入额外的网络结构,增大了模型的复杂度。因此,近期不少研究转向基于特征空间的数据增强方式。

Bengio 等[42]认为卷积操作将图像的流形线性化为一个深度特征的欧几里得子空间,因此,可以通过对特征空间进行线性操作实现复杂的属性转换任务[43]。风格迁移研究[44]表明,DNNs 的深层特征的统计信息包含了图像风格信息。基于此发现,Zhou 等[45]提出了 MixStyle,通过混合基于特征空间提取的不同域的风格信息可以生成具有新风格的图像。Gong 等[46]发现训练期间在特征空间上嵌入高斯噪声能有效提升分类器的域泛化性能,由此设计了包含类别信息的协方差矩阵进行自适应的特征增强。此外,文献[47]通过识别源域中数据的主要变化模式,然后隐式包含沿这些方向的增强版本来执行特征增强。

3.1.4 基于频域空间的数据增广

基于傅立叶变换,Yang 等[48]提出了一种新的域适应方法 FDA。FDA 设计了一种简单的图像转换策略,通过交换源域和目标域的低频频谱来减少源域和目标域分布之间的差异。通过简单训练幅度转移的源图像,FDA 取得了显著的泛化性能。受 FDA 启发,文献[49]提出了一种基于傅里叶变换的数据增广方法,该方法的设计动机来源于傅立叶变换的一个众所周知的特性[50],即傅立叶频谱的相位分量保留了原始信号的高级语义,而幅度分量包含低级统计信息。因此,Xu 等[49]在保留频谱相位信息的前提下,通过 MixUp 方式混合不同域的频谱幅度信息,以实现数据增广的目的。该方法可以避免过度拟合幅度信息中携带的低级统计信息,从而使决策时更加关注与高级语义相关的相位信息。类似地,Huang 等[51]提出了频率空间域随机化算法,通过离散余弦变换(Discrete Cosine Transform,DCT)将图像映射到频域空间,然后使用带通滤波器将其分成 64 个频率分量,通过划分并保持域不变频率分量(Domain-Invariant Frequency components, DIFs)和随机改变域可变频率分量(Domain-Variant Frequency components, DVFs)来实现频域空间下的数据增广。

3.2 优化模型求解

域泛化研究中优化模型解空间的指导思想是降低域偏移现象对模型求解过程的影响,根据方法设计思想的不同可分成 3 类:①基于标准化(Normalization)优化求解;②基于集成优化求解;③基于元学习优化求解。

3.2.1 基于标准化优化求解

由于域泛化研究中域偏移情况的存在,不同域数据协变量偏移(Covariate shift)形式不同,导致域间的统计特征(均值、方差)存在差异,使得传统的标准化方法,如 BN(Batch Normalization)、LN(Layer Normalization)和 IN(Instance Normalization)泛化性能不佳。域偏移会导致模型求解困难,增加模型复杂度,进而降低泛化性。

IBNet[52]实验发现 IN 能够在有效保留图片内容的同时,将图片的风格信息过滤掉,因此,IBNet 将 IN 引入到网络低层中,过滤掉低层特征中的外观信息,在跨域语义分割任务上取得了大幅的性能提升。Chang 等[53]为每一个域的数据设计了专属的结合 IN 和 BN 的标准化层来获取域特异(Domain-specific)的统计信息。为了使 IN 能自适应不同的域,文献[54]提出了一种通用的自适应标准化方法 ASR-Norm,它使用自动编码器让网络自动学习不同域下的 IN 归一化参数和缩放参数。在文献[55]中,作者提出了一种新的标准化方法,从 2 个方面解决分布变化问题:扩大训练时的分布及缩小测试时的分布。类似于 MixStyle[45],该方法设计了 CrossNorm 在训练过程中交换不同通道或不同实例的归一化参数,以模拟不同的域风格,之后基于注意力机制设计了 SelfNorm 模块让网络学习标准化中的缩放参数。

3.2.2 基于集成优化求解

集成学习(Ensemble learning)算法的基本思想是利用集成的方式平滑模型解空间,从而避免模型陷入局部最优。对于域泛化,集成学习通过使用特定的网络结构设计和训练策略来利用多个源域之间的关系,从而提高泛化性。

(1)模型集成

为每一个源域设计特定网络结构的模型集成方法[56],是域泛化中基于集成学习优化求解最直接的方式。为了进一步优化,Xu 等[57]认为浅层网络主要用于提取一般化特征,不同模型之间可以共享浅层网络参数,从而降低模型集成的计算开销。对于模型集成如何获得最终的预测,文献[58]采取直接对模型输出求平均的方法;Mancini 等[59]设计了域预测器用于预测样本属于每个域的概率,推断时用域预测器的概率作为权重,对不同源特定分类器的预测结果进行加权。在视网膜分割任务中,Wang 等[60]沿用了此策略用于集成预测。

(2)梯度集成

域泛化中集成学习的第二种方式是梯度集成。通常认为,基于模型集成的方法主要有 2 个缺陷:①模型

复杂度高;②难以捕捉不同域之间的关系。不同于模型集成每个域单独优化一个特定模型,基于梯度集成的方法同时利用多个源域的梯度信息共同优化一个模型。梯度集成最直接的方式是使用标准的小批量梯度下降(Mini-batch gradient descent),其中,mini-batch 是通过从所有源域中随机采样图像来构建的。Mansilla 等[61]认为每一个域内的数据包含特定于该域而与其他域无关的梯度信息,如果不加处理,域之间的梯度不一致会影响模型的泛化能力。他们设计了基于梯度符号一致性的判断策略,在优化过程中只对源域图像梯度一致的模型参数进行更新。Shi 等[62]认为源域梯度的方向和内积对模型学习域不变特征具有关键意义,因此提出域间梯度匹配算法,最大化梯度内积(GIP)以对齐跨域的梯度方向,并通过实验验证了梯度对齐在域泛化领域的有效性。

3.2.3　基于元学习优化求解

元学习,也叫学会学习,探索如何在训练任务中找到一些共性(Meta knowledge),作为先验知识帮助以后快速学习新的任务。基于梯度的元学习方法(Mode-Agnostic Meta-Learning, MAML)[20]最早应用于小样本学习。随后,Li 等[21]将 MAML 的情景训练范式(Episodic training paradigm)引入解决域泛化问题,提出 MLDG(Meta-Learning Domain Generalization)将多个源域随机划分成元训练域(Meta-train domain)和元测试域(Meta-test domain),并按照 MAML 方式进行训练。文章中证明了先在元训练域上更新一步,然后再在元测试域上更新的方式,相当于隐式地在不同域的更新梯度方向添加了一致性约束,这与前面基于梯度集成的方法思想是类似的,因此能提升域泛化能力。Balaji 等[63]认为 MLDG 可能不太适合目标数据不可见的场景,并且求二阶导操作对内存的消耗使其无法适用于大型网络。为解决上述问题,他们提出 MetaReg 算法,显式地学习只应用于网络分类层的正则化函数以提升模型泛化性。文献[64]指出上述研究都忽略了来自特征空间的语义信息指导,通过全局类对齐和局部样本聚类显式地约束了特征空间中的语义结构。

3.3　减小域间差异

第 2.2 小节理论上给出了跨域泛化性研究中影响目标域泛化误差上界的 3 个因素,分别是源域误差、理想联合预测误差和域间差异。因此,在模型给定的情况下,减小源域和目标域在假设函数下的分布差异是提升跨域泛化性的重要手段。在域泛化研究任务中,减小域间差异的方法大致分为 3 类:域不变特征表达、特征解耦以及一般化特征学习。

3.3.1　域不变特征表达

域不变性保证了特征对域偏移不敏感,因此能更好地泛化到不可见的目标域上。此类方法的关键在于寻找使不同域在映射空间内距离最小的映射函数,现有方法大致分为以下 3 类:

(1)基于核方法学习域不变特征

核方法(Kernel-based method)[65-68]是机器学习中最经典的学习方法之一。基于核方法可以将原始数据在高维映射空间中距离度量简单化,而无须关心映射函数的具体形式。域泛化期望在高维映射空间 $\varphi(\cdot)$ 内不同域的特征与标签的联合分布基本一致,从而令模型能学习域无关的特征,即

$$p^{(i)}(\varphi(X),Y)=p^{(j)}(\varphi(X),Y) \tag{14}$$

假设标注 Y 的条件分布 $p(Y|\varphi(X))$ 不随域发生改变,联合分布可以简化为源域和目标域的边缘分布一致。Pan 等[69]提出迁移成分分析方法(Transfer Component Analysis,TCA),将源域和目标域的数据映射到高维的再生核希尔伯特空间(Reproducing Kernel Hilbert Space, RKHS)。在 RKHS 空间中,最小化源域和目标域的最大均值差异(Maximum Mean Discrepancy,MMD),同时最大程度地保留域各自的内部属性。与 TCA 的核心思想相似,域不变成分分析法(Domain Invariant Component Analysis, DICA)[70]利用核方法对域泛化进行求解,目标是找到特征转换核 $k(\cdot,\cdot)$ 使所有数据在特征空间中的分布差异最小化。

实际场景下,不同域标注 Y 的条件分布一致的条件很难满足,依据贝叶斯公式

$$p(Y|\varphi(X))=p(\varphi(X)|Y)\cdot p(Y)/p(\varphi(X)) \tag{15}$$

令来自不同域的类别条件分布一致,即

$$p^{(i)}(\varphi(X)|Y)=p^{(j)}(\varphi(X)|Y) \tag{16}$$

可以通过对类别进行分组,约束相同类别不同域的特征尽量相似,并且要求不同类别的特征差异尽可能大来达到此目的。Ghifary 等[71]提出统一的框架 SCA(Scatter Component Analysis)对上述约束进行细化,并添加了在所有域上类内和类间不一致性的约束。而基于类别条件分布一致的在随后的域泛化研究[72-73]中得到了广泛关注。

(2)基于神经网络学习域无关特征

神经网络良好的非线性保证了其特征提取器可以起到核方法类似的效果,基于神经网络的方法优势在于能自动学习数据的映射方式。其目标函数与基于核方法的函数类似,要求不同域相同类别的特征距离尽量小,不同类别的特征尽量疏远。度量特征分布距离方式

常见的有基于统计的一阶、二阶矩(均值、方差)[74]、KL散度(Kullback-Leibler divergence)[75-76]、对比损失(Contrastive loss)[77-78]、最大均值差异MMD[69-71]以及三元组损失(Triplet loss)[64]等。在文献[76]中,KL散度被用来约束所有源域特征服从高斯分布。Dou等[64]提出了一种度量学习方案,通过一个度量学习网络ψ来约束不同域之间相同类别的特征尽可能紧凑,并设计了triplet loss进行距离计算。

(3)基于对抗学习域无关特征

和显式地度量分布距离不同,域对抗学习提供了一种隐式地学习域不变特征的方法。简单来说,在模型上设计一个域鉴别器,优化的目标是使不同域数据在特征空间上的分布无法分辨。Ganin等[79]首先在域自适应任务中提出了域对抗神经网络(Domain Adversarial Neural Network, DANN),通过交替对抗训练分类网络和域鉴别网络,使模型无法判断输入数据来自于源域还是目标域,从而达到学习域无关特征的目的。为了使得源域间特征之间类别条件分布一致如式(16),除了全局的域鉴别器,文献[80]还设计了多个不同类别先验下的条件域鉴别器。图像分类器和域鉴别器以逆梯度的方法(Reverse gradient)进行对抗训练,期望特征提取模型在训练的过程中能混肴全局域鉴别器。值得注意的是,条件域鉴别器与全局域鉴别器不同在于输入的样本属于同一个类别但不同域的样本数据。Li等[81]将MMD距离与自编解码(Autoencoder)结构结合提出了MMD-AAE框架,最小化自编码特征的域间MMD距离使得模型学到合适的全局特征。同时,作者认为在最优条件下,来自不同域的特征均值向量应当服从正态分布。基于该假设,模型设计了一个对抗结构,约束特征均值向量逼近正态分布生成的向量,MMD-AAE整体结构如图4所示。DLOW[82]在嵌入对抗的框架里利用对抗损失作为分布距离度量控制中间域与源域和目标域的相关性,从而去学习中间最优的特征变换。

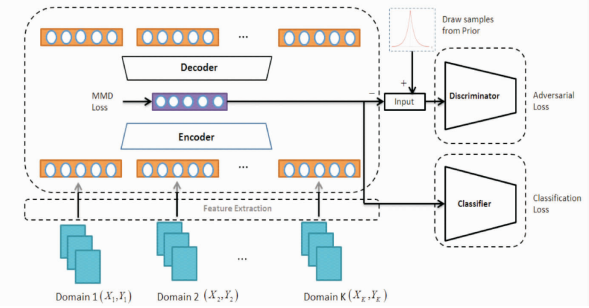

图4 MMD-AAE框架示意图

Fig. 4 Framework of MMD-AAE

3.3.2 特征解耦

当域偏移情况严重时,约束整个特征空间具有域不变性面临巨大的挑战。基于特征解耦的域泛化方法认为特征空间可以分解为子特征的组合,其中一部分属于域共享的域不变特征,另一部分是私有的域特异特征,从而可以利用特征解耦减弱特征域特异部分对域泛化的影响。

特征解耦主要有2种方式,一种从模型层面将参数分解成2部分:一部分负责提取域不变特征,另一部分提取域特异特征。文献[83]提出Undo-Bias将基于SVMs的分类器参数按此分成2部分,且仅用域不变特征处理未知域。同样地,神经网络模型也可以按参数进行分解,在文献[84]中,作者对Undo-Bias进行了扩展并设计了一个低秩参数化的CNN模型用于端到端域泛化学习。Chattopadhyay等[85]为了学习域特异和域不变特征之间的平衡,引入了域特异的激活掩码(Domain mask),使模型能够受益于域特异特征的预测能力,同时保持域不变特征的泛化性。Piratla等[86]直接对模型的权重矩阵应用低秩分解,以识别更通用的共同特征。另一种特征解耦的方式需要借助于生成模型。Ilse等[87]提出了域不变变分自编码器(Domain Invariant Variational Autoencoder, DIVA),DIVA是一个生成模型,通过学习3个独立的潜在子空间(类别、域和数据本身)解决域泛化问题。在文献[88]中DAL(Domain Agnostic Learning)特征被解耦为互信息最小化约束下的域不变特征、域特异特征和类无关特征3个部分(图5),DAL结合对抗训练和变分自编码器对原始特征进行重建,进而学习到域不变的特征表示。

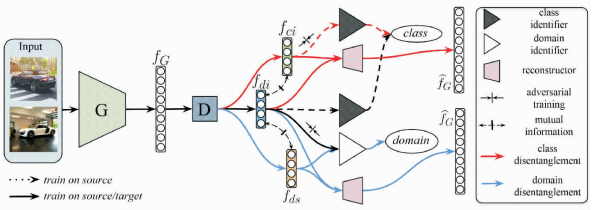

图5 特征解耦方法DAL示意图

Fig. 5 Illustration of feature disentanglement method DAL

3.3.3 一般化特征学习

一般化的特征学习主要包含2类方式,第一种是通过多任务学习的方式使其能同时处理多个子任务,从而优化原特征表达,增强特征的一般性。Ghifary等[89]设计了一种多任务的降噪自编码器(Multi-Task Autoencoder, MTAE),与传统自编码器从噪声中重建图像不同,MTAE学习利用原始图像同时重建多个相关域的图像。

因此,它可以学习对跨域变化更具有鲁棒性的特征。第二种方式是基于自监督学习(Self-supervised)思想,构造经验式(Empirical)或是启发式(Heuristic)的辅助自监督任务,通过辅助任务(Pretext)的学习,从无监督数据中挖掘数据自身的有用信息,从而提取更一般化的数据特征。常见的辅助任务有拼图游戏[90](Jigsaw puzzles)、预测变换参数[91]等。Jigsaw puzzles 构造了一个自监督的拼图任务供模型进行学习,并认为与域无关的任务可以减小域之间的分布差异。一般化特征学习的方法需要更多的工作以衡量其在域泛化研究中的效果。

4　应用场景与数据集

深度学习技术和计算资源算力的飞速发展,推动着计算机视觉领域中许多研究任务从实验室走向线下应用,如图像分类、动作识别、语义分割和行人重识别等。考虑到模型训练样本的有限性以及实际应用场景中数据的多样性,域泛化研究对于提升模型的泛化能力具有重要意义,目前,域泛化技术已在多种应用场景下开展广泛研究。

4.1　图像分类

4.1.1　数字识别

作为计算机视觉领域研究非常活跃且极具挑战性的应用场景之一,数字识别研究不仅具有丰富的理论价值,同时也具有很高的应用价值,在大规模数据统计和财务税务等金融领域具有十分广阔的应用前景。数字识别应用要求识别系统具有高识别精度和可靠性,然而,数字虽然笔画简单,却带有明显的个人特性,同一数字写法千差万别,就算是印刷体也会受字体不同的影响,这为域泛化研究提供了广阔的舞台。

目前,域泛化研究数字识别数据集主要包含 Digits-DG[92],Colored MNIST[93] 和 Rotated MNIST[89]。其中,Digits-DG 数据集由 MNIST[94]、MNIST-M[79]、SVHN[95] 和 SYN[79]4 种数据集构成。其中,MNIST 是手写数字图像数据集;MNIST-M 由 MNIST 和 BSDS500[96] 数据集中的随机色块混合而成;SVHN 是真实场景中的门牌号码图像;SYN 是基于 Windows™ 字体生成的合成数据集。这 4 个数据集由于在字体风格、笔画颜色和背景上存在明显的域差异,被用于数字识别域泛化性研究。Colored MNIST 和 Rotated MNIST 是在 MNIST 上分别将颜色和旋转角度作为域偏移量进行构建的,也是较为常用的域泛化数字识别数据集。

4.1.2　图像分类

图像分类是计算机视觉领域的基本问题,也是其他高层视觉任务(如目标检测、语义分割)的研究基础,在人脸识别,图像分类归档,智慧交通等领域具有广泛的应用价值。受图像生成方式、采集设备差异以及天气等的影响,图像数据集中域偏移现象较为普遍,这对模型域泛化能力提出了更高的要求,也成为域泛化研究最为热门的任务场景。常用的数据集介绍如下:

(1) CIFAR-10-C、CIFAR-100-C 和 ImageNet-C[97] 是分别在 CIFAR-10、CIFAR-100[98] 和 ImageNet[99] 数据集的基础上,进行了不同类型的破坏,如添加噪声、模糊、天气和数字化。

(2) PACS[6]、Office-Home[100]、DomainNet[101] 和 ImageNet-R[102] 这 4 个数据集均关注图像不同的风格变化。其中,PACS 由 4 种风格的图像组成,包括美术绘画、卡通、照片和素描;Office-Home 中的图像风格包括艺术、剪纸、产品和现实世界;DomainNet 包含 6 种不同的风格;而最新提出的 ImageNet-R[102] 包含了原始 ImageNet 中 200 个目标类别的各种艺术再现。

(3) VLCS[103]、Office-31[104] 和 Terr Incognita[105] 主要关注不同环境的域差异。VLCS 由 Caltech101[106]、PASCAL VOC[107]、LabelMe[108] 和 SUN09[109] 这 4 个不同的数据集组合而成,包含不同环境和视角的变化;Office-31 数据集由常见的办公设备组成,包含 3 个域的目标(Amazon、DSLR 和 WebCam);Terr Incognita 由 4 个不同地域的野生动物图像构成。

4.2　语义分割

语义分割是图像理解中的关键一环,在自动驾驶和地理信息系统等领域具有广阔的应用前景。语义分割中域泛化研究主要围绕传感器采集条件不同,如天气、光照和季节的影响导致的图像域偏移问题展开。由于数据的稀缺性和医疗成像设备的差异性,域泛化在医学影像分析领域也具有重要的应用价值。考虑到数据获取和标注的难度,语义分割数据集主要由真实场景数据集和合成场景数据集组成。

真实场景数据集包括 Cityscapes[110]、BDD-100K[111] 和 Mapillary[112]。Cityscapes 是包含 50 个不同城市高分辨率场景图像的大规模数据集;BDD-100K 源于美国不同地方采集的驾驶图像;Mapillary 包含 25 000 张从世界各地采集的高分辨率街景图像。合成数据集包括 GTA5[113] 和 SYNTHIA[114] 数据集。GTA5 数据集由游戏生成,从汽车视角拍摄,包含 24 966 张具有像素级语义标注的合成图像,与真实场景数据集共享 19 个类别;SYNTHIA 数据集包括 3 个地点,以及不同的天气、光照和季节,与真实场景数据集共享 16 个类别。

此外,在医学图像分割任务上,域泛化研究往往在

不同模态的数据以及不同医疗机构获取的数据上开展研究。

4.3 安防监控

安防监控领域的应用主要包括人脸活体检测、深度伪造检测和行人重识别等场景。此类应用场景涉及到严重的信任问题和社会安全,对模型的泛化能力具有更高的要求。图像攻击方法和篡改技术的多样性,为安防监控领域算法的域泛化性研究提供了基础。

4.3.1 人脸活体检测

近年来,诸如打印攻击、视频攻击、3D 掩码攻击等攻击方法的出现,给人脸识别技术的广泛应用带来巨大安全风险。由于攻击类型的多样化,以及显示设备的不同,在进行人脸活体检测时,提升模型域泛化能力至关重要。此任务场景中常用的是 COMI[115],由 OULUN-PU[116]、CASIA-FASD[117]、Idiap Replay-Attack[118] 和 MSU-MFSD[119] 4 种攻击方法生成的图像构成。人脸活体检测域泛化研究可以有效提升检测系统的安全性和鲁棒性。

4.3.2 深度伪造检测

随着人脸合成技术的发展,深度伪造图像和视频越来越逼真。由于伪造算法的多样性,开发鲁棒的深度伪造检测模型十分关键。FaceForensics ++[120] 是研究深度伪造检测域泛化任务的常用数据集。使用 4 种最先进的人脸伪造方法:Deepfake[121]、Face2Face[122]、FaceSwap[123] 和 NeuralTextures[124] 构建。深度伪造检测域泛化研究遵循 leave-one-domain-out[6] 评价规则;或者使用 FaceForensics ++ 中的所有数据训练,然后在其他数据集,如 DeeperForensics[125]、FaceShifter[126]、Celeb-DF-v2[127] 和 DFDC[128] 上进行测试[129]。

4.3.3 行人重识别

行人重识别域泛化问题中域偏移通常来源于不同的相机、视图、光照和背景等条件。常用的数据集包含 VIPeR[130]、PRID[131]、CUHK02[132]、CUHK03[133]、Duke[134]、Market[135] 和 MSMT17[136],主要通过相机来区分不同域。此外,最新发布的 Person30K[137] 数据集解决了以往数据集样本和标注身份有限、采样相机少、环境条件和姿势变化少等问题,提出了一个更大规模的行人重识别数据集,以帮助提升行人重识别模型的表示和泛化能力。

表 2 总结了当前与域泛化研究的相关应用场景以及常用数据集和数据集基本信息。

表 2 域泛化应用场景与数据集

Table 2 Applications and popular datasets for domain generalization

任务	数据集	域	样本数	描述
数字识别	Digits-DG[92]	4	24 000	来自 4 个数据集(MNIST、MNIST-M、SVHN 和 SYN)
	Colored MNIST[93]	3	70 000	不同域的红色或绿色手写数字
	Rotated MNIST[89]	6	70 000	6 种不同旋转角度的手写数字
目标识别	CIFAR-10-C[97]	–	60 000	不同类型的破坏,如噪声、模糊、天气和数字化,每种破坏有 5 个等级
	CIFAR-100-C[97]	–	60 000	
	ImageNet-C[97]	–	≈1.3M	
	ImageNet-R[102]	–	30 000	原始 ImageNet 中 200 个目标类别的艺术再现
	PACS[6]	4	9 991	图像风格不同(美术绘画、卡通、照片和素描)
	Office-Home[100]	4	15 588	图像风格不同(艺术、剪纸、产品和现实世界)
	DomainNet[101]	6	586 575	图像风格不同(剪贴画、表意、速写、绘画、真实和素描)
	VLCS[103]	4	10 729	来自 Caltech101、PASCAL VOC、LabelMe 和 SUN09
	Office-31[104]	3	4 652	Amazon、DSLR 和 WebCam 的办公设备
	Terr Incognita[105]	4	24 788	不同地点的野生动物
语义分割	GTA5[113]	–	24 966	合成数据集
	SYNTHIA[114]	–	2 700	合成数据集,包括 3 个地点,以及不同的天气、光照和季节
	Cityscapes[110]	–	5 000	50 个不同城市的高分辨率城市场景图像
	BDD-100K[111]	–	8 000	从美国不同地方收集的驾驶图像
	Mapillary[112]	–	25 000	从世界各地采集的高分辨率街景图像
人脸活体检测	COMI[115]	4	8 500	来自 4 个数据集(OULUNPU、CASIA-FASD、Idiap Replay-Attack 和 MSU-MFSD)

(续表 2)

任务	数据集	域	样本数	描述
深度伪造检测	FaceForensics ++[120]	4	5 000	4 种不同人脸伪造方法(Deepfake、Face2Face、FaceSwap 和 NeuralTextures)生成
行人重识别	VIPeR[130]	–	1 264	2 个不同的相机
	PRID[131]	–	1 134	2 个不同的相机
	CUHK02[132]	–	7 267	2 个不同的相机
	CUHK03[133]	–	28 192	2 个不同的相机
	Duke[134]	–	36 411	8 个不同的相机
	Market[135]	–	32 668	6 个不同的相机
	MSMT17[136]	–	126 411	15 个不同的相机,2 种场景
	Person30K[137]	–	1 384 940	6 497 个不同的相机,89 个拍摄地点,4 个季节,2 种场景

5　讨论与展望

近些年来,尽管一系列致力于提升模型在未知域下泛化性的域泛化技术被提出,并取得了一定的成果,但是域偏移问题还远没有被解决,域泛化研究仍然充满挑战。在人工智能应用落地速度加快的大背景下,真实应用场景下数据的多样性和不可预知性对模型泛化性提出了巨大挑战,域泛化问题仍然是一个需要深耕的方向。本节简要介绍域泛化领域未来的一些迫切需要研究的方向。

(1)异质化的域泛化研究。当前研究基本集中在同质化的域泛化研究,即目标域与源域的标签空间是一致的。然而实际场景下目标域数据可能会是源域内没有出现过的类别,这要求模型拥有类似零样本学习的能力。另外,源域内域偏移的形式也是同质化的,当前域泛化研究往往只关注单一域偏移形式,如 PACS 数据集只关注图像风格的不同、VLCS 只关注环境和视角的变化。现实中的目标域偏移具有不可预知性,可能与模型训练时的域偏移情况相差甚远,从而导致模型的性能大大下降。这要求域泛化研究需要同时关注多种不同形式的域偏移。

(2)增量式域泛化研究。模型部署后参数通常是固定不变的,但目标域数据的分布是不停变化着的,当数据的分布发生剧烈变化时模型可能会失效。因此,希望域泛化研究拥有增量学习的能力,能够不断地处理现实世界中连续的信息流,在吸收新知识的同时保留甚至整合、优化旧知识的能力,避免灾难性遗忘。

(3)探索新的数据合成方式。数据增广的方式在域泛化研究领域是最直接且有效的方式,丰富源域数据的多样性对于域泛化研究至关重要。然而有些应用收集数据是极其困难的,如医学影像数据;同时,在某些任务中数据标注成本高昂,比如语义分割。数据合成方式提供了一种经济可行的方式,目前已有一些合成数据的方法,但是面临计算开销大,多样性有限等问题。因此,探索新的数据合成算法具有现实意义。

(4)半监督域泛化研究。当前域泛化研究默认源域数据都是有标注的,且受限于标注代价,数据集样本数明显少于监督学习中的设定。现实中获取大量的未标注数据相对来说是更容易的,半监督的域泛化值得被研究。半监督学习中常用策略,如伪标签生成、一致性正则对于域泛化研究是否依然奏效,半监督信息和监督信息对于模型优化来说是否可以当成不同的域进行处理。这些问题都需要被研究后才会有答案。

(5)联邦域泛化研究。现有的域泛化研究需要在学习过程中访问多源分布,然而出于数据隐私考虑,具有分布式数据源的联邦范式给域泛化研究带来了新的挑战。在联邦范式中,数据是分布存储的,每个客户端的学习都只能访问本地数据,因此,当前的域泛化研究方法在联邦范式下是不适用的。另外,局部优化将使模型倾向于自身的数据分布,难以推广到新的目标域。因此,联邦范式下的域泛化研究也可能成为未来的一个研究方向。

6　总　结

现实场景中数据的多变性和不可预知性对模型的域泛化能力提出巨大的挑战,研究域泛化技术对于模型部署应用具有重要意义。本文梳理总结了近年来计算机视觉领域内的域泛化研究工作,对域泛化的任务定义、任务特点和研究思路做了详细的概述。并遵循域泛化研究思路,将域泛化研究现有方法分成 3 大类,并阐

述了每个类别下具有代表性的技术和典型算法。此外,介绍了目前域泛化技术在计算机视觉领域中的应用场景和公开数据集。最后,本文对域泛化领域值得进一步深入探索的方向进行了展望,指出了当前域泛化面临的问题与挑战。

参考文献:

[1] 王郁夫,李沛辰,易波,等. 人工智能赋能网络安全应用[J]. 广州大学学报(自然科学版), 2021, 20(2):1-12.

[2] 尹奇跃,黄岩,张俊格,等. 基于深度学习的跨模态检索综述[J]. 中国图象图形学报, 2021, 26(6):1368-1388.

[3] Recht B, Roelofs R, Schmidt L, et al. Do ImageNet classifiers generalize to ImageNet? [C]//The 36th International Conference on Machine Learning. Long Beach: PMLR, 2019: 5389-5400.

[4] Hendrycks D, Dietterich T. Benchmarking neural network robustness to common corruptions and perturbations[EB/OL]. (2019-05-28)[2022-03-22]. https://arxiv.org/pdf/1903.12261.pdf.

[5] Blanchard G, Lee G, Scott C. Generalizing from several related classification tasks to a new unlabeled sample[C]//The 25th Conference on Neural Information Processing Systems. Granada: NIPS, 2011: 2178-2186.

[6] Li D, Yang Y, Song Y Z, Hospedales T M. Deeper, broader and artier domain generalization[C]//The 16th International Conference on Computer Vision. Piscataway: IEEE, 2017:5542-5550.

[7] Zhang X, Cui P, Xu R, et al. Deep stable learning for out-of-distribution generalization[C]//The 34th IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021:5372-5382.

[8] Shu Y, Cao Z, Wang C, et al. Open domain generalization with domain-augmented meta-learning[C]//The 34th IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021:9624-9633.

[9] Wang J, Lan C, Liu C, et al. Generalizing to unseen domains: A survey on domain generalization[C]//The 30th International Joint Conference on Artificial Intelligence. Virtual: Ijcai.org, 2021:4627-4635.

[10] Zhou K, Liu Z, Qiao Y, et al. Domain generalization: A survey[EB/OL]. (2021-07-18)[2022-04-22]. https://arxiv.org/pdf/2103.02503.pdf.

[11] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[C]//The 27th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 1701-1708.

[12] 翁金塔,仇晶,张光华. 面向推理的知识图谱表示学习方法综述[J]. 广州大学学报(自然科学版), 2021, 20(3): 80-89.

[13] Geirhos R, Temme C R M, Rauber J, et al. Generalisation in humans and deep neural networks[C]//The 32nd Conference on Neural Information Processing Systems. Montreal: NIPS, 2018:7549-7561.

[14] Hoffman J, Tzeng E, Park T, et al. Cycada: Cycle-consistent adversarial domain adaptation[C]//The 35th International Conference on Machine Learning. Stockholm: PMLR, 2018:1994-2003.

[15] Mallya A, Davis D, Lazebnik S. Piggyback: Adapting a single network to multiple tasks by learning to mask weights[C]//The 15th European Conference on Computer Vision. Berlin: Springer, 2018: 67-82.

[16] Guo P, Lee C Y, Ulbricht D. Learning to branch for multi-task learning[C]//The 37th International Conference on Machine Learning. Virtual: PMLR, 2020: 3854-3863.

[17] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 22(10): 1345-1359.

[18] Weiss K, Khoshgoftaar T M, Wang D D. A survey of transfer learning[J]. Journal of Big Data, 2016, 3(1): 1-40.

[19] Zhuang F, Qi Z, Duan K, et al. A comprehensive survey on transfer learning[J]. Proceedings of the IEEE, 2020, 109(1): 43-76.

[20] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//The 34th International Conference on Machine Learning. Sydney: PMLR, 2017:1126-1135.

[21] Li D, Yang Y, Song Y Z, et al. Learning to generalize: Meta-learning for domain generalization[C]//The 32nd AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2018: 3490-3497.

[22] Wang W, Zheng V W, Yu H, et al. A survey of zero-shot learning: Settings, methods, and applications[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-37.

[23] Long M, Cao Y, Wang J, et al. Learning transferable features with deep adaptation networks[C]//The 32nd International Conference on Machine Learning. Lille: PMLR, 2015: 97-105.

[24] Liu Z, Miao Z, Pan X, et al. Open compound domain adaptation[C]//The 33rd IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020:12406-12415.

[25] 李航. 统计学习方法[M]. 北京:清华大学出版社,2012.

[26] Ben-David S, Blitzer J, Crammer K, et al. Analysis of representations for domain adaptation[C]//The 20th Conference on Neural Information Processing Systems. Vancouver: NIPS, 2006: 133-144.

[27] Ben-David S, Blitzer J, Crammer K, et al. A theory of learning from different domains[J]. Machine Learning, 2010, 79(1): 151-175.

[28] Zhang C, Bengio S, Hardt M, et al. Understanding deep learning requires rethinking generalization[EB/OL]. (2017-02-26)[2022-03-22]. https://arxiv.org/pdf/1611.03530.pdf.

[29] Krueger D, Ballas N, Jastrzebski S, et al. Deep nets don't learn via memorization[EB/OL]. (2017-02-21)[2022-03-22]. https://openreview.net/pdf?id=rJv6ZgHYg.

[30] Arpit D, Jastrzębski S, Ballas N, et al. A closer look at memorization in deep networks[C]//The 34th International Conference on Machine Learning. Sydney: PMLR, 2017:233-242.

[31] Volpi R, Murino V. Addressing model vulnerability to distributional shifts over image transformation sets[C]//The 17th International Conference on Computer Vision. Piscataway: IEEE, 2019: 7980-7989.

[32] Shi Y, Yu X, Sohn K, et al. Towards universal representation learning for deep face recognition[C]//The 33rd Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 6817-6826.

[33] Yue X, Zhang Y, Zhao S, et al. Domain randomization and pyramid consistency: Simulation-to-real generalization without accessing target domain data[C]//The 17th International Conference on Computer Vision. Piscataway: IEEE, 2019: 2100-2110.

[34] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//The 16th International Conference on Computer Vision. Piscataway: IEEE, 2017: 2223-2232.

[35] Rahman M M, Fookes C, Baktashmotlagh M, et al. Multi-component image translation for deep domain generalization[C]//IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2019:579-588.

[36] Anoosheh A, Agustsson E, Timofte R, et al. ComboGAN: Unrestrained scalability for image domain translation.[C]//The 31st Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2018: 783-790.

[37] Gretton A, Borgwardt K M, Rasch M J, et al. A kernel two-sample test[J]. The Journal of Machine Learning Research, 2012, 13(1): 723-773.

[38] Volpi R, Namkoong H, Sener O, et al. Generalizing to unseen domains via adversarial data augmentation[C]//The 32rd Conference on Neural Information Processing Systems. Montreal: NIPS, 2018: 5339-5349.

[39] Qiao F, Zhao L, Peng X. Learning to learn single domain generalization[C]//The 33rd Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 12556-12565.

[40] Shankar S, Piratla V, Chakrabarti S, et al. Generalizing across domains via cross-gradient training[EB/OL]. (2018-05-01)[2022-03-22]. https://arxiv.org/pdf/1804.10745.pdf.

[41] Zhou K, Yang Y, Hospedales T, et al. Deep domain-adversarial image generation for domain generalization[C]//The 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2020: 13025-13032.

[42] Bengio Y, Mesnil G, Dauphin Y, et al. Better mixing via deep representations[C]//The 30th International Conference on Machine Learning. Atlanta: PMLR, 2013: 552-560.

[43] Upchurch P, Gardner J, Pleiss G, et al. Deep feature interpolation for image content changes[C]//The 30th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 7064-7073.

[44] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization[C]//The 16th International Conference on Computer Vision. Piscataway: IEEE, 2017: 1501-1510.

[45] Zhou K, Yang Y, Qiao Y, et al. Domain generalization with mixstyle[EB/OL]. (2021-04-05)[2022-03-22]. https://arxiv.org/pdf/2104.02008.pdf.

[46] Gong Y, Lin X, Yao Y, et al. Confidence calibration for domain generalization under covariate shift[C]//The 18th Interna-

tional Conference on Computer Vision. Piscataway: IEEE, 2021: 8958-8967.

[47] Khan M H, Zaidi T, Khan S, et al. Mode-guided feature augmentation for domain generalization[C]//The 32nd British Machine Vision Conference. Virtual:BMVC, 2021.

[48] Yang Y, Soatto S. Fda: Fourier domain adaptation for semantic segmentation[C]//The 33rd Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 4085-4095.

[49] Xu Q, Zhang R, Zhang Y, et al. A fourier-based framework for domain generalization[C]//The 34th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 14383-14392.

[50] Wang H, Wu X, Huang Z, et al. High-frequency component helps explain the generalization of convolutional neural networks[C]//The 33rd Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 8684-8694.

[51] Huang J, Guan D, Xiao A, et al. Fsdr: Frequency space domain randomization for domain generalization[C]//The 34th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 6891-6902.

[52] Pan X, Luo P, Shi J, et al. Two at once: Enhancing learning and generalization capacities via IBN-Net[C]//The 15th European Conference on Computer Vision. Berlin: Springer, 2018: 464-479.

[53] Chang W G, You T, Seo S, et al. Domain-specific batch normalization for unsupervised domain adaptation[C]//The 32nd Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 7354-7362.

[54] Fan X, Wang Q, Ke J, et al. Adversarially adaptive normalization for single domain generalization[C]//The 34th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 8208-8217.

[55] Tang Z, Gao Y, Zhu Y, et al. CrossNorm and selfNorm for generalization under distribution shifts[C]//The 18th International Conference on Computer Vision. Piscataway: IEEE, 2021: 52-61.

[56] Zhou Z H. Ensemble methods: Foundations and algorithms[M]. Boca Raton: CRC Press, 2012.

[57] Xu H, Xie H, Liu Y, et al. Deep cascaded attention network for multi-task brain tumor segmentation[C]//The 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2019: 420-428.

[58] D'Innocente A, Caputo B. Domain generalization with domain-specific aggregation modules[C]//The 40th German Conference on Pattern Recognition. Berlin: Springer, 2018: 187-198.

[59] Mancini M, Bulo S R, Caputo B, et al. Best sources forward: Domain generalization through source-specific nets[C]//The 25th International Conference on Image Processing. Piscataway: IEEE, 2018: 1353-1357.

[60] Wang S, Yu L, Li K, et al. Dofe: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets[J]. IEEE Transactions on Medical Imaging, 2020, 39(12): 4237-4248.

[61] Mansilla L, Echeveste R, Milone D H, et al. Domain generalization via gradient surgery[C]//The 18th International Conference on Computer Vision. Piscataway: IEEE, 2021: 6630-6638.

[62] Shi Y, Seely J, Torr P H S, et al. Gradient matching for domain generalization[EB/OL]. (2021-07-14)[2022-04-22]. https://arxiv.org/pdf/2104.09937.pdf.

[63] Balaji Y, Sankaranarayanan S, Chellappa R. Metareg: Towards domain generalization using meta-regularization[C]//The 32nd Conference on Neural Information Processing Systems. Montreal: NIPS, 2018: 1006-1016.

[64] Dou Q, Coelho de Castro D, Kamnitsas K, et al. Domain generalization via model-agnostic learning of semantic features[C]//The 33rd Conference on Neural Information Processing Systems. Vancouver: NIPS, 2019: 6447-6458.

[65] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.

[66] Blanchard G, Lee G, Scott C. Generalizing from several related classification tasks to a new unlabeled sample[C]//The 25th Conference on Neural Information Processing Systems. Granada: NIPS, 2011: 2178-2186.

[67] Long M, Wang J, Ding G, et al. Transfer feature learning with joint distribution adaptation[C]//The 14th International Conference on Computer Vision. Piscataway: IEEE, 2013: 2200-2207.

[68] Wang J, Chen Y, Hao S, et al. Balanced distribution adaptation for transfer learning[C]//The 17th International Conference on Data Mining. Piscataway: IEEE, 2017: 1129-1134.

[69] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. IEEE Transactions on Neural Networks, 2010, 22(2): 199-210.

[70] Muandet K, Balduzzi D, Schölkopf B. Domain generalization via invariant feature representation[C]//The 14th International Conference on Machine Learning. Atlanta: PMLR, 2013: 10-18.

[71] Ghifary M, Balduzzi D, Kleijn W B, et al. Scatter component analysis: A unified framework for domain adaptation and domain generalization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(7): 1414-1430.

[72] Hu S, Zhang K, Chen Z, et al. Domain generalization via multidomain discriminant analysis[C]//The 35th Conference on Uncertainty in Artificial Intelligence. Tel Aviv: AUAI, 2020: 292-302.

[73] Li Y, Gong M, Tian X, et al. Domain generalization via conditional invariant representations[C]//The 32nd AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2018: 3579-3587.

[74] Erfani S, Baktashmotlagh M, Moshtaghi M, et al. Robust domain generalisation by enforcing distribution invariance[C]//The 25th International Joint Conference on Artificial Intelligence. Palo Alto: AAAI, 2016: 1455-1461.

[75] Wang Z, Loog M, Gemert J V. Respecting domain relations: Hypothesis invariance for domain generalization[C]//The 26th International Conference on Pattern Recognition. Piscataway: IEEE, 2021: 9756-9763.

[76] Li H, Wang Y F, Wan R, et al. Domain generalization for medical imaging classification with linear-dependency regularization[C]//The 34th Conference on Neural Information Processing Systems. Virtual: NIPS, 2020: 3118-3129.

[77] Motiian S, Piccirilli M, Adjeroh D A, et al. Unified deep supervised domain adaptation and generalization[C]//The 19th International Conference on Computer Vision. Piscataway: IEEE, 2017: 5715-5725.

[78] Yoon C, Hamarneh G, Garbi R. Generalizable feature learning in the presence of data bias and domain class imbalance with application to skin lesion classification[C]//The 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2019: 365-373.

[79] Ganin Y, Lempitsky V. Unsupervised domain adaptation by Backpropagation[C]//The 32nd International Conference on Machine Learning. Lille: PMLR, 2015: 1180-1189.

[80] Li Y, Tian X, Gong M, et al. Deep domain generalization via conditional invariant adversarial networks[C]//The 15th European Conference on Computer Vision. Berlin: Springer, 2018: 624-639.

[81] Li H, Pan S J, Wang S, et al. Domain generalization with adversarial feature learning[C]//The 31st Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 5400-5409.

[82] Gong R, Li W, Chen Y, et al. Dlow: Domain flow for adaptation and generalization[C]//The 32nd Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 2477-2486.

[83] Khosla A, Zhou T, Malisiewicz T, et al. Undoing the damage of dataset bias[C]//The 12nd European Conference on Computer Vision. Berlin: Springer, 2012: 158-171.

[84] Li D, Yang Y, Song Y Z, et al. Deeper, broader and artier domain generalization[C]//The 16th International Conference on Computer Vision. Piscataway: IEEE, 2017: 5542-5550.

[85] Chattopadhyay P, Balaji Y, Hoffman J. Learning to balance specificity and invariance for in and out of domain generalization[C]//The 16th European Conference on Computer Vision. Berlin: Springer, 2020: 301-318.

[86] Piratla V, Netrapalli P, Sarawagi S. Efficient domain generalization via common-specific low-rank decomposition[C]//The 37th International Conference on Machine Learning. Virtual: PMLR, 2020: 7728-7738.

[87] Ilse M, Tomczak J M, Louizos C, et al. Diva: Domain invariant variational autoencoders[C]//Medical Imaging with Deep Learning. Montreal: PMLR, 2020: 322-348.

[88] Peng X, Huang Z, Sun X, et al. Domain agnostic learning with disentangled representations[C]//The 36th International Conference on Machine Learning. Long Beach: PMLR, 2019: 5102-5112.

[89] Ghifary M, Kleijn W B, Zhang M, et al. Domain generalization for object recognition with multi-task autoencoders[C]//The 15th International Conference on Computer Vision. Piscataway: IEEE, 2015: 2551-2559.

[90] Noroozi M, Favaro P. Unsupervised learning of visual representations by solving Jigsaw puzzles[C]//The 14th European Conference on Computer Vision. Berlin: Springer, 2016: 69-84.

[91] Zhang L, Qi G J, Wang L, et al. Aet vs. aed: Unsupervised representation learning by auto-encoding transformations rather than data[C]//The 32nd Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 2547-2555.

[92] Zhou K, Yang Y, Hospedales T, et al. Learning to generate novel domains for domain generalization[C]//The 16th European Conference on Computer Vision. Berlin: Springer, 2020: 561-578.

[93] Nam J, Cha H, Ahn S, et al. Learning from failure: De-biasing classifier from biased classifier[C]//The 34th Conference on Neural Information Processing Systems. Virtual: NIPS, 2020: 20673-20684.

[94] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[95] Netzer Y, Wang T, Coates A, et al. Reading digits in natural images with unsupervised feature learning[C]//The 25th NIPS Workshop on Deep Learning and Unsupervised Feature Learning. Granada: NIPS, 2011.

[96] Arbelaez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(5): 898-916.

[97] Hendrycks D, Dietterich T. Benchmarking neural network robustness to common corruptions and perturbations[EB/OL]. (2019-03-28)[2022-03-22]. https://arxiv.org/pdf/1903.12261.pdf.

[98] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[EB/OL]. (2009-04-08)[2022-04-22]. http://citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.222.9220&rep=rep1&type=pdf.

[99] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[100] Venkateswara H, Eusebio J, Chakraborty S, et al. Deep hashing network for unsupervised domain adaptation[C]//The 30th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 5018-5027.

[101] Peng X, Bai Q, Xia X, et al. Moment matching for multi-source domain adaptation[C]//The 32nd International Conference on Computer Vision. Piscataway: IEEE, 2019: 1406-1415.

[102] Hendrycks D, Basart S, Mu N, et al. The many faces of robustness: A critical analysis of out-of-distribution generalization [C]//The 18th International Conference on Computer Vision. Piscataway: IEEE, 2021: 8340-8349.

[103] Fang C, Xu Y, Rockmore D N. Unbiased metric learning: On the utilization of multiple datasets and web images for softening bias[C]//The 14th International Conference on Computer Vision. Piscataway: IEEE, 2013: 1657-1664.

[104] Saenko K, Kulis B, Fritz M, et al. Adapting visual category models to new domains[C]//The 11st European Conference on Computer Vision. Berlin: Springer, 2010: 213-226.

[105] Beery S, Horn G V, Perona P. Recognition in terra incognita[C]//The 15th European conference on computer vision. Berlin: Springer, 2018: 456-473.

[106] Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories[C]//CVPR Workshops. Piscataway: IEEE, 2004: 178-178.

[107] Everingham M, Gool L V, Williams C K I, et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[108] Russell B C, Torralba A, Murphy K P, et al. LabelMe: A database and web-based tool for image annotation[J]. International Journal of Computer Vision, 2008, 77(1): 157-173.

[109] Choi M J, Lim J J, Torralba A, et al. Exploiting hierarchical context on a large database of object categories[C]//The 23rd Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2010: 129-136.

[110] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]//The 29th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 3213-3223.

[111] Yu F, Chen H, Wang X, et al. Bdd100k: A diverse driving dataset for heterogeneous multitask learning[C]//The 33rd Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 2636-2645.

[112] Neuhold G, Ollmann T, Bulo S R, et al. The mapillary vistas dataset for semantic understanding of street scenes[C]//The 16th International Conference on Computer Vision. Piscataway: IEEE, 2017: 4990-4999.

[113] Richter S R, Vineet V, Roth S, et al. Playing for data: Ground truth from computer games[C]//The 14th European Conference on Computer Vision. Berlin: Springer, 2016: 102-118.

[114] Ros G, Sellart L, Materzynska J, et al. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes[C]//The 29th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 3234-3243.

[115] Quan R, Wu Y, Yu X, et al. Progressive transfer learning for face anti-spoofing[J]. IEEE Transactions on Image Processing, 2021, 30: 3946-3955.

[116] Boulkenafet Z, Komulainen J, Li L, et al. OULU-NPU: A mobile face presentation attack database with real-world variations[C]//The 12th International Conference on Automatic Face & Gesture Recognition. Piscataway: IEEE, 2017: 612-

618.

[117]Zhang Z, Yan J, Liu S, et al. A face antispoofing database with diverse attacks[C]//The 5th International Conference on Biometrics. Piscataway: IEEE, 2012: 26-31.

[118]Chingovska I, Anjos A, Marcel S. On the effectiveness of local binary patterns in face anti-spoofing[C]//International Conference of Biometrics Special Interest Group. Piscataway: IEEE, 2012: 1-7.

[119]Wen D, Han H, Jain A K. Face spoof detection with image distortion analysis[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(4): 746-761.

[120]Rossler A, Cozzolino D, Verdoliva L, et al. Faceforensics + +: Learning to detect manipulated facial images[C]//The 17th International Conference on Computer Vision. Piscataway: IEEE, 2019: 1-11.

[121]DeepFakes[CP/OL]. [2022-03-29]. http://www.github.com/deepfakes/faceswap.

[122]Thies J, Zollhofer M, Stamminger M, et al. Face2face: Real-time face capture and reenactment of RGB videos[C]//The 29th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 2387-2395.

[123]FaceSwap[CP/OL]. [2022-03-29]. http://www.github.com/MarekKowalski/FaceSwap.

[124]Thies J, Zollhöfer M, Nießner M. Deferred neural rendering: Image synthesis using neural textures[J]. ACM Transactions on Graphics, 2019, 38(4): 1-12.

[125]Jiang L, Li R, Wu W, et al. Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection[C]//The 33rd Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2020: 2889-2898.

[126]Li L, Bao J, Yang H, et al. Faceshifter: Towards high fidelity and occlusion aware face swapping[EB/OL]. (2020-09-15)[2020-04-22]. https://arxiv.org/pdf/1912.13457.pdf.

[127]Li Y, Yang X, Sun P, et al. Celeb-df: A large-scale challenging dataset for deepfake forensics[C]//The 33rd Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 3207-3216.

[128]Dolhansky B, Bitton J, Pflaum B, et al. The deepfake detection challenge (DFDC) dataset[EB/OL]. (2020-10-28)[2022-03-22]. https://arxiv.org/pdf/2006.07397.pdf.

[129]Haliassos A, Vougioukas K, Petridis S, et al. Lips don't lie: A generalisable and robust approach to face forgery detection[C]//The 34th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 5039-5049.

[130]Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]//The 10th European Conference on Computer Vision. Berlin: Springer, 2008: 262-275.

[131]Hirzer M, Beleznai C, Roth P M, et al. Person re-identification by descriptive and discriminative classification[C]//Scandinavian Conference on Image Analysis. Berlin: Springer, 2011: 91-102.

[132]Li W, Wang X. Locally aligned feature transforms across views[C]//The 26th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE 2013: 3594-3601.

[133]Li W, Zhao R, Xiao T, et al. Deepreid: Deep filter pairing neural network for person re-identification[C]//The 27th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 152-159.

[134]Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking[C]//ECCV Workshops. Berlin: Springer, 2016: 17-35.

[135]Zheng L, Shen L, Tian L, et al. Scalable person re-identification: A benchmark[C]//The 15th International Conference on Computer Vision. Piscataway: IEEE, 2015: 1116-1124.

[136]Wei L, Zhang S, Gao W, et al. Person transfer gan to bridge domain gap for person re-identification[C]//The 31st Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 79-88.

[137]Bai Y, Jiao J, Ce W, et al. Person30k: A dual-meta generalization network for person re-identification[C]//The 34th Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 2123-2132.

【责任编辑: 卓祯雨】